

# Measurement system analysis (MSA) of empirically-derived composite measure of preventive care counseling practices of HIV medical care providers

Osaro Mgbere<sup>1,2</sup>, Mamta Singh<sup>1</sup>, Raouf Arafat<sup>1</sup>

<sup>1</sup>Bureau of Epidemiology, Office of Surveillance and Public Health Preparedness, Houston Department of Health and Human Services, Houston, Texas, USA.

<sup>2</sup>Institute of Community Health, University of Houston College of Pharmacy, Texas Medical Center, Houston, Texas, USA.  
Correspondence to: Osaro Mgbere, E-mail: Osaro.Mgbere@Houstontx.gov

Received March 15, 2015. Accepted May 28, 2015

## Abstract

**Background:** Although preventive care services and risk reduction counseling are offered to HIV patients during their routine clinic visits, assessment of clinicians practice performance has been difficult because of the use of many separate indicators.

**Objective:** To assess the measurement system of empirically-derived composite measure of preventive care practices of HIV medical care providers (overall preventive care counseling index, OPCi) in managed care clinical settings using the following statistical properties: precision (gauge reproducibility and repeatability, gauge R&R), accuracy (bias and linearity), stability, and process capability.

**Materials and Methods:** Data used were obtained from the cross-sectional survey of HIV medical care providers in 13 HIV care facilities in Houston/Harris County, Texas, USA, that participated in the Centers for Disease Control and Prevention (CDC) Medical Monitoring Project in 2009.

**Result:** Domain-specific preventive care counseling indices developed produced significant ( $p \leq 0.05$ ) Cronbach's alpha coefficients that ranged from 0.64 to 0.91. The variance components for OPCi attributed to the provider and patient statuses were 29.7% and 31.4%, respectively, while the gauge R&R was 68.6%. A high percent gauge R&R precision to tolerance ratio (99.13%) and precision to total variation ratio (82.85%) were obtained. With measurement system intraclass correlations without bias factors ( $r_{pe}$ ) of 0.45 and with bias factors ( $r_b$ ) of 0.31, the system was classified as "third class monitor," indicating the need for process improvement. However, the system expanded uncertainty was within set range at 99% confidence level making it statistically "stable" and "capable" at a process capability ratio ( $C_p$ ) of 0.999 (95% CI: 0.793–1.205) and long-term *sigma* yield of 2.292.

**Conclusion:** Reduction in measurement variations would enhance providers' performance output and improve both the quality of preventive care counseling received and the overall health outcome of HIV patients.

**KEY WORDS:** HIV/AIDS, HIV medical care provider, preventive care counseling, composite index, measurement system analysis

## Introduction

Until early 2000s, the HIV-infected population has been understudied and underserved with respect to risk reduction and prevention interventions because of the concentration of efforts on at-risk HIV-negative populations. Preliminary studies have shown increases in high-risk sex practices and sexually transmitted diseases (STDs) among HIV-positive individuals receiving care.<sup>[1-4]</sup> These findings suggest that

### Access this article online

Website: <http://www.ijmsph.com>

DOI: 10.5455/ijmsph.2015.15032015336

Quick Response Code:



reduction in HIV transmission in the United States will require new strategies and led to the establishment of guidelines for incorporating HIV prevention into the medical care of persons living with HIV<sup>[5]</sup> along with new HIV prevention strategies and programs.<sup>[6-8]</sup> Brief prevention messages delivered by HIV medical care providers (HMCPs) can change patients' behaviors in ways that decrease their risk for transmitting HIV. Although preventive care services and risk reduction counseling are offered to HIV patients during their routine clinic visits, assessment of clinicians practice performance has been difficult because of the use of many separate indicators and differing methods of documentation. In addition, inconsistent measures and lack of alignment constitute a major obstacle in the effort to improve preventive care for patients.

Recently, there has been more interest in public reporting and pay for performance (P4P), and with the passage of the Affordable Care Act (ACA) in 2010, health and healthcare in the United States are in the midst of unprecedented change. Although P4P incentives have traditionally been aimed at rewarding preventive service delivery in primary care practices,<sup>[9]</sup> under the ACA, it replaces traditional fee-for-service, and thus, measuring the quality of care is a critical starting point for improvement in providers' performance over time. In this changing health-care landscape, quality of care offered by providers will have important implications for new payment models, care coordination, health-care costs and may even be the key determinant of how long someone with HIV lives.

The quality measures used in P4P typically fall into four categories, namely: structure, process, outcome, and patient experience measures.<sup>[9]</sup> Combining measures of performance not only convey a broader picture when compared with a single measure, but also hold promise for improving understanding and stimulating improvement among providers. Individual measures of performance that reflect the quality of care provided to patients by providers or institutions, when combined into a composite performance measure, would be useful in detecting a consistent pattern of practice or quality of care across patients of the provider or institution. On a larger scale, composite measures can facilitate system-level changes by highlighting the need for better care coordination and accountability across multiple providers.<sup>[10]</sup>

A composite measure has been defined as a combined metric that incorporates multiple individual measures to provide a single score.<sup>[11]</sup> A number of different approaches for developing a composite measure of provider performance have been studied.<sup>[11-16]</sup> The American Medical Association (AMA) through the Physician Consortium for Performance Improvement (PCPI)<sup>[11]</sup> and National Quality Forum (NQF)<sup>[12]</sup> have also developed policy framework for composite measures with greater consistency and standardization. The PCPI affirms that adding composite measures to its portfolio can provide an overall summary of the quality of care delivered to a patient, at both the clinician and the patient levels.

Although performance measurement has been in place for some time in hospitals and managed care organizations, the

focus on providers' profiling is a relatively new development. Public and private purchasers of insurance and health plans are now demanding more information about the quality and relative cost of US physicians to not only increase physician accountability, but also to aid in value-based purchasing.<sup>[17]</sup> The inherent limitations of the available data at the physician level have brought to the forefront technical issues such as performance measurement reliability, validity, and bias, which were less important at higher levels of aggregation in hospitals and managed care organizations.<sup>[17]</sup>

In an earlier study,<sup>[18]</sup> we assessed the frequency and determinants of preventive care counseling by HMCPs using empirically-derived composite indices and reported significant variations based on providers' practice characteristics in favor of newly diagnosed patients compared with the established ones. These observations were associated, possibly, with providers' performance measurement reliability, validity, and bias. As with individual measures of performance, the methods used to construct empirically-derived composites can affect the reliability, validity, and usefulness of the measures and require some unique considerations for testing and analysis.<sup>[12]</sup> This study, therefore, was aimed at evaluating the empirically-derived overall preventive care counseling index (OPCi) as a measure of HMCPs' performance in managed care clinical settings in Houston/Harris County, Texas. We conducted measurement system analysis of this measure of performance with the following specific objectives: (i) to determine the variability in the performance data collected that is due to the measurement system; (ii) to isolate the components of variability of the measurement system; and (iii) to assess the measurement system precision, accuracy, stability, and capability (and suitability for some improvement, if required).

## Materials and Methods

### Design, Participants, and Data Collection

The survey participants consisted of HMCPs from 13 HIV care facilities in Houston/Harris County, Texas, who participated in the Centers for Disease Control and Prevention (CDC)'s Medical Monitoring Project (MMP) Provider Survey. The survey was conducted in Houston/Harris County, Texas, from June through September 2009. HMCP eligible for this survey included physicians, physician assistants, or nurse practitioners at the MMP sampled facilities/providers who have provided care to, ordered CD4+ or HIV viral load testing for, and/or prescribed ART to HIV-infected adults  $\geq 18$  years of age. Physicians who were interns, residents, fellows, and others in training programs were not eligible. Eligible providers were required to independently assess their preventive care counseling practices for newly diagnosed and established HIV patients. A detailed description of the MMP providers' survey design and data collection instrument and procedures can be found in McNaghten *et al.*<sup>[19]</sup> and Mgbere *et al.*<sup>[20]</sup> The study sample in Houston/Harris County comprised of 51 HMCPs and recorded a response rate of 45%. The demographic and

medical practice characteristics of participating HIV care providers have also been described in detail previously.<sup>[20]</sup>

### Development of Preventive Care Counseling Indices

HMCPs were asked to indicate how often they counseled their patients using 20 identified preventive care measures. These measures cover items related to medication and adherence, risk reduction counseling, mental health and substance use, and disease screenings and support services. Using a scoring scheme that ranged from 4 (always discuss) to 1 (almost never discuss), 20 preventive care counseling measures were used to develop six composite counseling indices (CCIs) such that a higher score indicated the provider's likelihood to engage in preventive care counseling activity, while a lower score represented less likelihood to engage in preventive care counseling activity. The indicator variables were not transformed since they all had a common unit of measurement. However, to indicate their relative importance, differential weights based on conceptual rationale and priority of measures<sup>[21,22]</sup> were applied to maximize reliability and discrimination as follows: 0.250 for medication and adherence, risk reduction, mental health, and substance use indicators; and 0.125 for disease screenings and support services, with the weights summing up to 1.0.<sup>[18]</sup> The weights were then multiplied by the actual performance values for each indicator variable and then summed to produce the composite scores or indices.

The six composite preventive care counseling indices developed include the following: medication and adherence index (MAi) that used a 5-item scale that includes adherence with antiretroviral regimen, adherence with opportunistic infection (OI) prophylaxis, how to take medicines, medical-related side effects, and drug–drug interactions; risk reduction index (RRi), which measures HIV transmission risk-related counseling activities by providers, used 6-items, namely: risk reduction regarding HIV transmission, condom use, pregnancy or potential pregnancy with female patients, availability of partner counseling services, disclosure of HIV status to their partners, and pregnancy or potential pregnancy with female patients not on ART; disease screening index (DSi) was assessed using a 5-item scale that includes screening for sexually transmitted disease, hepatitis, tuberculosis, hepatitis (A and B) vaccination, and discussion about tuberculosis disease risk. Mental health and substance use index (MSi) was evaluated using a 2-item scale namely mental health problems, including depression and substance use; and social and family support index (SSi) comprised of 2 items, namely: wellness nutrition, exercise, etc.) and family/social support. OPCi was made up of the 5 subcomposite scales (MAi, RRi, DSi, MSi, and SSi), and comprised of 20 items. More information about the empirically-derived composite measures of performance and the associated procedures can also be found in Mgbere et al.<sup>[18]</sup> In its simplest form, the CCIs take a linear form as follows:

$$CCI_n = W_1 y_{1n} + W_2 y_{2n} + \dots + W_p y_{pn} = \sum_i W_i y_{in}$$

where  $CCI_n$  is the composite score for unit  $n$  (MAi, RRi, DSi, MSi, SSi, and OPCi);  $y_{in}$  is the individual performance measure for attribute  $i$  in unit  $n$ , and  $w_i$  is the weight attached to attribute  $i$ .<sup>[21]</sup>

### Measurement System and Statistical Analysis

The measurement systems analysis (MSA)<sup>[23]</sup> was used to assess the precision, consistency, and bias of our measurement system when it interacts with its environment. In addition, MSA allowed us to classify the measurement system including the responsiveness of process control efforts and determine the positive or negative impact of any improvement strategy. The multi-item six composite preventive care counseling indices developed were first tested for internal consistency reliability using Cronbach's alpha coefficient. However, we used only the OPCi in our current study to determine the provider performance because the measure included various aspects of the counseling process. Using the means chart, we compared the individual provider group mean values with the overall mean for OPCi. The analysis of variance and variance components analysis were further used to model the variation from measurement to measurement with various levels of classification using the following independent factors: provider, patient, and interaction between provider and patient statuses.

### The Gauge R&R Study

To properly monitor and improve on the measurement system used, it was necessary to conduct a gauge reproducibility and repeatability (gauge R&R) study to measure variability and other attributes of the process output. The gauge R&R analysis method<sup>[24–29]</sup> was adapted and used in the current study to determine how much of the variability in our measurement system was attributed to provider variation (reproducibility), measurement variation (repeatability), and process variation using the crossed effects model. Simply put, the gauge R&R is the variance equal to the sum of within-system and between-system variations. The components of the measurement system sources of variation are presented in Figure 1, while the statistical model for gauge R&R in our study was expressed as a two-way random effect:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, i = 1, 2, \dots, I, j = 1, 2, \dots, J, k = 1, 2, \dots, r.$$

where: the observation,  $Y_{ijk}$  is the  $k$ th assessment of provider  $j$  on  $i$  patient status;

$\mu$  is the unknown grand mean.

$\alpha_i \sim N(0, \sigma_{\alpha})$ , which are the random effects of different parts (patient status);

$\beta_j \sim N(0, \sigma_{\beta})$ , which are the random effects of different providers;

$\alpha\beta_{ij} \sim N(0, \sigma_{\alpha\beta})$ , which are the random effects of patient status–provider interaction;

$e_{ijk} \sim N(0, \sigma)$ , which is the random error due to measurement replications.

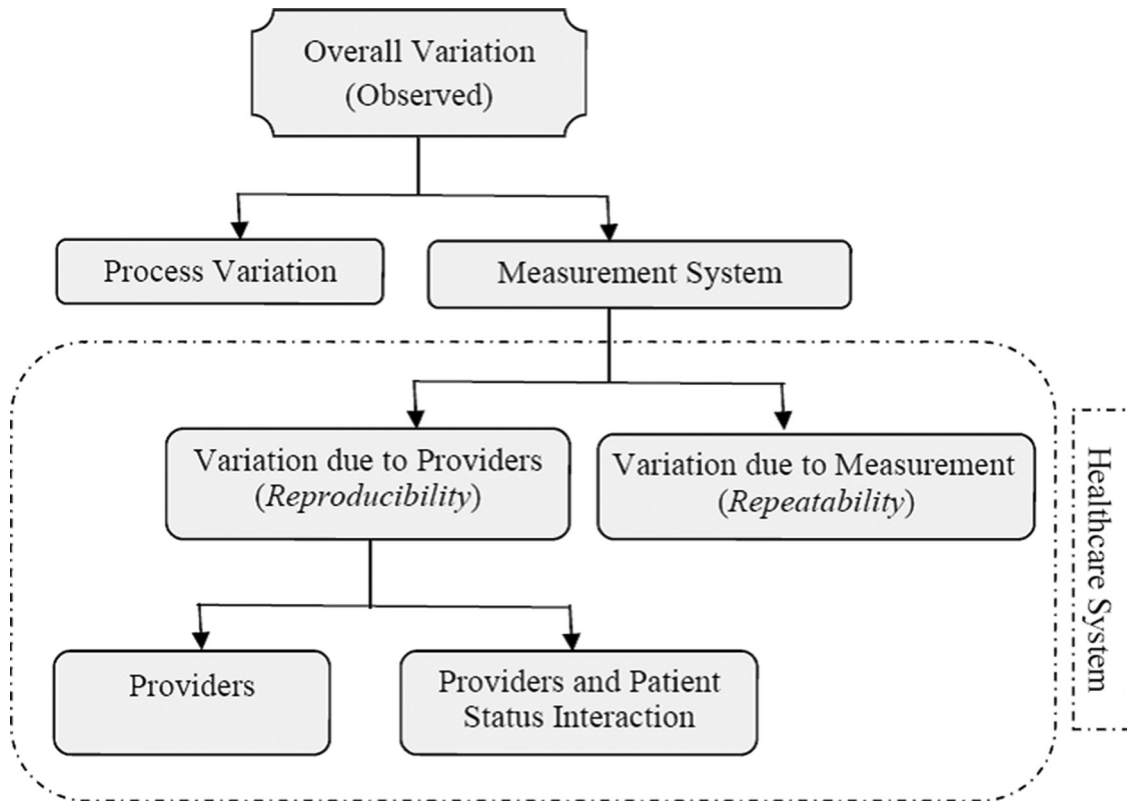


Figure 1: Components of preventive care counseling measurement sources of variation.

There are four independent variance components:  $\sigma_a^2, \sigma_\beta^2, \sigma_{a\beta}^2, \sigma^2$ . If the estimated variance component became negative, the value was set to zero.

Conceptually,  $\sigma_{\text{observed process}}^2 = \sigma_{\text{actual/process}}^2 + \sigma_{\text{measurement system}}^2$

$$\sigma_{\text{measurement system}}^2 = \sigma_{\text{repeatability}}^2 + \sigma_{\text{reproducibility}}^2$$

$$\sigma_{\text{repeatability}} = \sigma$$

$$\sigma_{\text{reproducibility}} = \sqrt{\sigma_\beta^2 + \sigma_{a\beta}^2}$$

$$\begin{aligned} \sigma_{\text{Gauge R\&R}} &= \sqrt{\sigma_\beta^2 + \sigma_{a\beta}^2 + \sigma^2} \\ &= \sqrt{\sigma_{\text{reproducibility}}^2 + \sigma_{\text{repeatability}}^2} \end{aligned}$$

### Measurement Precision/Tolerance

It was assumed that our process exhibited statistical control and involved a quality characteristic that followed a normal distribution with mean,  $\mu$  and standard deviation,  $\sigma$ , where the upper and lower natural tolerance limits of the process were  $\mu + 3\sigma$  and  $\mu - 3\sigma$ . The measurement system (gauge R&R) precision was determined using the following measures: Precision-to-part (patient status) variation (PPV) was defined as:

$$PPV = 6 \times \sigma_{RR} / 6 \times \sigma_{PV}$$

Percent precision to tolerance ratio (PTR) was defined as the ratio between total measurement error (gauge R&R) and the part tolerance;

$$PTR = 100 \times \frac{6 \times \sigma_{RR}}{USL - LSL}$$

where: USL and LSL are the upper and lower specification limits, respectively;  $\sigma_{RR}$  = Gauge R&R standard error.

Percent precision to total variation ratio:

$$PTVR = 100 \times \frac{6 \times \sigma_{RR}}{6 \times \sigma_{TV}}$$

The number of distinct categories (NDC), which represents the number of groups that our measurement system can distinguish from the data, was assessed as:

$$NDC = [1.41(6 \times \sigma_{PV} / 6 \times \sigma_{RR})]$$

and rounded down to the nearest integer. Six is used as the multiplier ( $k$ ) in these formulas, because for the normal distribution, it covers 99.73% of the process. Furthermore, the discrimination ratio (DR) was used to measure the relative

usefulness of a specific factor by comparing the total variance of the measurement with the variance of the measurement error. The DR was computed for the main effects as follows:

$$DR = \sqrt{2 \left( \frac{P}{TV - P} \right) + 1}$$

where:  $P$  = estimated variance for a factor (patient status, provider);  $TV$  = estimated total variance.

**Effective Resolution**

The effective resolution (ER) measures were used to assess how well our measurement increments (MI) are working and determine the need to retain, add, or drop digits when recording the measurements.

Probable error (PE) was computed as follows:

$$Z_{0.75} \times \sigma_{pe}$$

where:

$\hat{\sigma}_{pe}^2$  = variance estimate for within (error);

$Z_{0.75}$  = the 75% quantile of standard normal distribution

Lower bound increment was calculated as: 0.1PE;  
 Smallest effective increment was calculated as: 0.22PE;  
 Largest effective increment was calculated as: 2.2PE.

**Intraclass Correlation**

The similarity of observations within patient status (groups) relative to that among the groups were evaluated using intraclass correlation coefficient ( $r$ ). In the context of MSA, the following types of intraclass correlations were determined:

Intraclass correlation without bias was computed as follows:

$$r_{pe} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{pe}^2}$$

Intraclass correlation with bias was computed as follows:

$$r_b = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_b^2 + \hat{\sigma}_{pe}^2}$$

Intraclass correlation with bias and interaction factors was computed as follows:

$$r_{int} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_b^2 + \hat{\sigma}_{int}^2 + \hat{\sigma}_{pe}^2}$$

where:

$\hat{\sigma}_{pe}^2$  = variance estimate for within (error);

$\hat{\sigma}_p^2$  = variance estimate for product;

$\hat{\sigma}_b^2$  = variance estimate for bias factors;

$\hat{\sigma}_{int}^2$  = variance estimate for interaction factors.

The likelihood of detecting a shift in the measurement system process was approximated using the four classes of process monitors (first, second, third, and fourth)<sup>[28,29]</sup> based on the estimated intraclass correlation coefficients. The amount by which a signal coming from the measurement process is attenuated by the effects of measurement error was estimated by:

$$\text{Process signal attenuation} = 1 - \sqrt{\text{intraclass correlation}}$$

**Performance Distribution Pattern**

The normal quantile plot was used to visualize the extent to which OPCi was normally distributed with the normal quantile plot showing the confidence bounds, probability, and normal quantile scales.

The normal quantile values were computed as follows:

$$\Phi^{-1} \left( \frac{r_i}{N+1} \right)$$

While the empirical cumulative probability for each value was computed as:

$$\frac{r_i}{N+1}$$

where:

$\Phi$  is the cumulative probability distribution function for the normal distribution;

$r_i$  is the rank of the  $i$ th observation;

$N$  is the number of nonmissing observations.

Cumulative distribution function (CDF) was used to determine the probability of providers attaining a given performance output based on the OPCi score and represented as:

$$F(x) = \Phi \left( \frac{x - \mu}{\sigma} \right)$$

**Process Capability Indices**

The ability of the measurement system to track process improvements was quantified by using process capabilities determined as the ratio of the spread between the process specifications (the specification “width”) to the spread of the process values, as measured by six process standard deviation units (the process width). All capability analyses use the same general formulas but differ in how sigma ( $\sigma$ ) is computed, which are as follows:

$$\text{Process capability ratio } (C_p) = \frac{(USL - LSL)}{6\sigma}$$

The lower and upper confidence intervals (CIs) for  $C_p$  were computed with the formulas:

$$\text{Lower CI for } C_p: C_p \sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}}$$

$$\text{Upper CI for } C_p: C_p \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}}$$

Process capability index,  $C_{pk} = \min \left[ \frac{(USL - \mu)}{3 \times \sigma}, \frac{(\mu - LSL)}{3 \times \sigma} \right]$

CIs for  $C_{pk}$ :

$$\text{Lower CI: } \hat{C}_{pk} \left[ 1 - \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{1}{9n\hat{C}_{pk}^2} + \frac{1}{2(n-1)}} \right]$$

$$\text{Upper CI: } \hat{C}_{pk} \left[ 1 + \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{1}{9n\hat{C}_{pk}^2} + \frac{1}{2(n-1)}} \right]$$

Process capability index,

$$C_{pm} = \frac{\min(\text{target} - LSL, USL - \text{target})}{3\sqrt{\sigma^2 + (\text{mean} - \text{target})^2}} = \frac{(USL - LSL)}{6\sqrt{\sigma^2 + (\mu - T)^2}}$$

where:  $T$  is the target mean.

CIs for  $C_{pm}$ :

$$\text{Lower CI: } C_{pm} \sqrt{\frac{\chi_{\alpha/2, \gamma}^2}{\gamma}}$$

$$\text{where } \gamma = \frac{n \left( 1 + \left( \frac{\bar{x} - \text{Target}}{s} \right)^2 \right)}{1 + 2 \left( \frac{\bar{x} - \text{Target}}{s} \right)^2}$$

$s$  = standard deviation

$$\text{Upper CI: } C_{pm} \sqrt{\frac{\chi_{1-\alpha/2, \gamma}^2}{\gamma}}$$

where  $\gamma$  = same as above.

Process capability ratio of one-sided upper spec,

$$C_{pu} = \frac{\text{Allowable upper spread}}{\text{Actual upper spread}} = \frac{(USL - \mu)}{3 \times \sigma}$$

Process capability ratio of one-sided lower spec,

$$C_{pl} = \frac{\text{Allowable lower spread}}{\text{Actual lower spread}} = \frac{(\mu - LSL)}{3 \times \sigma}$$

where  $\mu$  and  $\sigma$  are the process mean and standard deviation, respectively. Estimators of  $C_{pu}$  and  $C_{pl}$  are obtained by replacing  $\mu$  and  $\sigma$  by mean and  $\sigma$ , respectively. The following relationship holds:

$$C_p = \frac{(C_{pu} + C_{pl})}{2}$$

**Goal Plot**

To show how the performance measure (OPCi) is conforming to specification limits, the goal plot based on the spec-normalized mean shift on the x-axis and the spec-normalized

standard deviation on the y-axis was used. For each column with LSL, target, and USL, these quantities were defined as:

$$\text{Mean shift normalized to spec} = \frac{(\text{Mean}(\text{Col } [i]) - \text{Target})}{(USL [i] - LSL [i])}$$

$$\text{Standard deviation normalized to spec} = \frac{(\text{Std.Dev.}(\text{Col } [i]))}{(USL [i] - LSL [i])}$$

with three shaded regions representing the relationship between our system process capability ( $C_{pk} = 0.999, \approx 1.00$ ), which approximates a nonconformance rate of 2,719.78 part per million (ppm), if distribution is centered, and the resultant outcome from a shift in  $C_{pk}$  (Figures 5 and 6).

**Measurement Sigma**

Long-term sigma (overall),

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Short-term sigma,

$$\sigma = \sqrt{\frac{1}{r} \sum_{i=1}^r \sum_{j=1}^{n_j} \frac{(x_{ij} - \bar{x}_i)^2}{n_j - 1}}$$

where  $\bar{x}_i = \frac{1}{n_j} \sum_{j=1}^{n_j} x_{ij}$ ,  $r$  is the number of subgroups of size  $n_j$ ,

and each  $i$ th subgroup is defined by the order of the data.

**Sigma quality:**

$$\text{Sigma quality} = \text{Normal quantile} \left( 1 - \frac{\% \text{ outside}}{100} \right) + 1.5$$

$$\text{Sigma quality above} = \text{Normal quantile} \left( 1 - \frac{\% \text{ above}}{100} \right) + 1.5$$

$$\text{Sigma quality below} = \text{Normal quantile} \left( 1 - \frac{\% \text{ below}}{100} \right) + 1.5$$

For all tests performed, the probability value of 0.05 was used as threshold for determining statistical significance level. All measurement system and statistical analyses were performed using JMP statistical discovery™ software, version 11.1 (SAS Institute, Cary, NC, USA).

**Ethics Statement**

MMP was determined by the National Center for HIV, Viral Hepatitis, STD and TB Prevention's Office of the Associate Director for Science at the CDC to be a nonresearch, public health surveillance activity used for disease control program or policy purposes. As such, MMP is not subject to human subjects regulations, including federal investigational review board (IRB) review. As an amendment to MMP, the MMP

Provider Survey is covered under the same nonresearch determination. No local IRB was required for the Houston project area.

**Results**

Summary of preventive care counseling domain-specific internal consistency reliability estimates are presented in Table 1. The Cronbach’s alpha coefficient for the overall scale and subscales ranged from 0.64 (SSi) to 0.91 (DSi) indicating that the items that made up the domains had good internal consistency and, thus, measure the same concept or construct intended. Analysis of OPCi means and standard deviation by the participating HMCPs’ group is presented in Figure 2. The distribution represents the means performance variations of participating providers group and compares those with the overall mean outcome. The standard deviations allow for comparison of levels of bias across the providers performance scores.

Table 2 presents the MSA for OPCi of HMCPs. The MSA results indicate that provider, patient status, and within-process error contributed 29.7%, 31.4%, and 38.9% of the total variance components, respectively. There was

no significant variation associated with interaction effect between provider and patient status. Our measurement system variation (gauge R&R) was 68.6%, with repeatability having a higher variability (38.9%) compared with reproducibility (29.7%). The measurement PPV ratio was 1.48, while the percent gauge R&R PTR and PTVR were 99.13% and 82.85%, respectively. The estimate for NDC in our system was 1.0, while the DR for both provider and patient status sources of variation were 1.36 and 1.38, respectively. The ERs measured by PE and MI were 1.153 and 0.001 respectively, with the largest effective measurement increment expected at error value below 2.5373. The measurement increment of 0.001 was below the lowest measurement increment bound and, therefore, it was recommended that it be adjusted to record fewer digits.

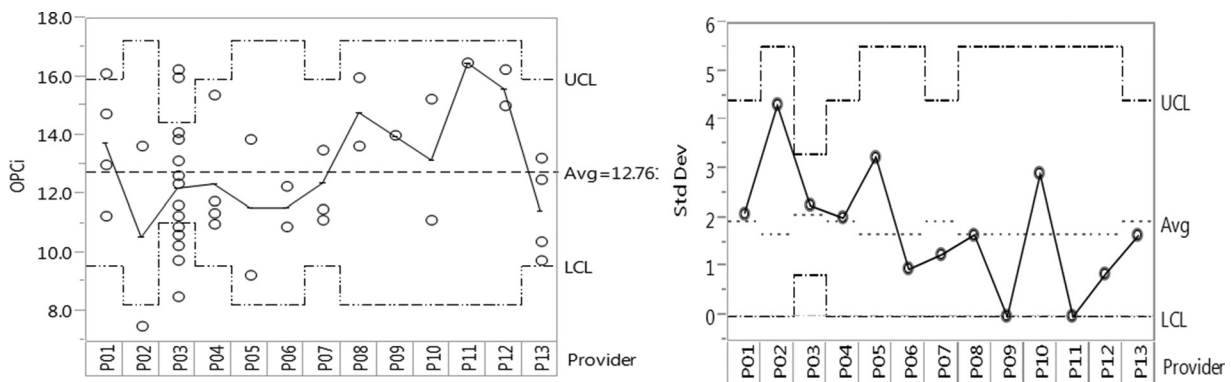
The intraclass correlation ( $r_{pe}$ ) representing the proportion of variation attributed to patient status without including bias factors was 0.446 and served as a reference point. However, including the bias factors reduced the impact by 29.7% to  $r_b = 0.314$ . Using the guidelines and standards for system monitor [Table 3],<sup>[27,29,30]</sup> our current measurement system was classified as third class monitor based on the intraclass correlation values with bias and without bias factors with attenuation of process signals recorded at 33.22% and

**Table 1:** Summary of preventive care counseling domain-specific internal consistency reliability estimates

Domain index	Items in scale (n)	Minimum score	Maximum score	Cronbach’s alpha	
				Coefficient	95% CI
OPCi <sup>a</sup>	20	7.50	16.50	0.79	0.68–0.87
MAi	5	2.75	5.00	0.88	0.82–0.93
RRi	6	2.25	6.00	0.90	0.84–0.94
DSi	5	0.75	2.50	0.91	0.86–0.94
MSi	2	0.50	2.00	0.80	0.67–0.88
SSi	2	0.38	1.00	0.64	0.41–0.79

OPCi, overall preventive care counseling index; MAi, medication and adherence index; RRi, risk reduction index; DSi, disease screenings; MSi, mental health and substance use index; SSi, social and family support index.

<sup>a</sup>This measure formed the basis for the measurement system analysis in this study.



**Figure 2:** Distribution of OPCi means and standard deviation by HIV medical care provider groups.

**Table 2:** Measurement system analysis for overall preventive care counseling index (OPCi) of HIV medical care providers

	Variance component	Standard deviation	% of Total	Sum of variances
<b>Source</b>				
Provider	2.23	1.49	29.7	$V(\text{provider})$
Patient status	2.36	1.53	31.4	$V(\text{patient status})$
Provider × patient status	0.00	0.00	0.00	$V(\text{provider} \times \text{patient status})$
Within	2.92	1.71	38.9	$V(\text{within})$
Total	7.51	2.74	100.0	$V(\text{within}) + V(\text{provider}) + V(\text{provider} \times \text{patient status}) + V(\text{patient status})$
<b>Measurement variation</b>				
Gauge R&R (RR)	5.16	2.27	68.6	$V(\text{within}) + V(\text{provider}) + V(\text{provider} \times \text{patient status})$
Repeatability (EV)	2.92	1.71	38.9	$V(\text{within})$
Reproducibility (AV)	2.23	1.49	29.7	$V(\text{provider}) + V(\text{provider} \times \text{patient status})$
Product variation (PV)	2.36	1.53	31.4	$V(\text{patient status})$
Interaction variation (IV)	0.00	0.00	0.00	$V(\text{provider} \times \text{patient status})$
Total variation (TV)	7.51	2.74	100.0	$V(\text{within}) + V(\text{provider}) + V(\text{provider} \times \text{patient status}) + V(\text{patient status})$
<b>Precision/tolerance</b>				
Precision to part variability ratio (PPV)	1.48			<b>Formula</b> $6 \times \sigma_{RR} / 6 \times \sigma_{PV}$
Tolerance (T)	13.74			USL-LSL <sup>a</sup>
% RR precision to tolerance ratio (PTR)	99.13			$\left[ 100 \times \frac{6 \times \sigma_{RR}}{(USL - LSL)} \right]$
% RR precision to total variation ratio (PTVR)	82.85			$[100 \times (6 \times \sigma_{RR} / 6 \times \sigma_{TV})]$
Number of distinct categories (NDC)	0.95≈1.00			$[1.41(6 \times \sigma_{PV} / 6 \times \sigma_{RR})]$
Discrimination ratio (DR)				$D = \sqrt{2 \left( \frac{P}{TV - P} \right) + 1}$
Provider	1.36			
Patient status	1.38			
<b>Effective resolution<sup>b</sup></b>				
Probable error (PE)			<b>Results</b>	<b>Value description</b>
			1.153	Median error for a single measurement $[0.675 \times \sigma_{\text{Repeatability}}]$
Current measurement increment (MI)			0.001	Measurement increment estimated from data (in tenths)
Lower bound increment (0.1 × PE)			0.1153	Measurement increment should not be below this value
Smallest effective increment (0.22 × PE)			0.2537	Measurement increment is more effective above this value
Largest effective increment (2.2 × PE)			2.5373	Measurement increment is more effective below this value

Measurement test <sup>c</sup>	System classification <sup>c</sup>	Results	Value description
Test-retest error		1.710	Within error
Intraclass correlation, $r_{pe}$ (no bias)	Third class	0.446	Proportion of variation attributed to patient status without including bias factors
Intraclass correlation, $r_b$ (with bias)	Third class	0.314	Proportion of variation attributed to patient status with bias factors
Intraclass correlation, $r_{int}$ (with bias and interactions)	Third class	0.314	Proportion of variation attributed to patient status with bias factors and interactions
Bias impact	—	0.133	Fraction by which the bias factors reduce the intraclass correlation
Bias and interaction impact	—	0.133	Fraction by which the bias factors and interactions reduce the intraclass correlation

<sup>a</sup>Upper specification limit (USL) = 19.63; lower specification limit (LSL) = 5.89.

<sup>b</sup>Action: recommends dropping a digit because the measurement increment of 0.001 is below the lowest measurement increment bound.

<sup>c</sup>Monitor classification legend<sup>[29,30]</sup>: refer to Table 3.

**Table 3:** Monitor classification legend<sup>[29,30]</sup>

Classification	Intraclass correlation	Attenuation of process signal (%)	Probability of warning, test 1 only <sup>a</sup>	Probability of warning, tests 2-4 only <sup>a</sup>
First class	0.80–1.00	Less than 11	0.99–1.00	1.0
Second class	0.50–0.80	11–29	0.88–0.99	1.0
Third class	0.20–0.50	29–55	0.40–0.88	0.92–1.00
Fourth class	0.02–0.20	More than 55	0.30–0.40	0.08–0.92

<sup>a</sup>Probability of warning for a 3 standard error shift within 10 subgroups using Wheeler's tests.

43.96%, respectively, and a chance of detecting a 3-standard error shift.

Table 4 shows the descriptive statistics, specification limits, prior probability (of control probability), and process sigma estimation of providers' preventive care counseling performance. With a normalized mean shift to specification of -0.272 and prior probability of 0.272, a nonconformance rate of 2,719.78 ppm at a sigma quality of 4.280 was obtained for total outside portion of the sample distribution [Table 4]. The long-term yield, representing the percentage of successful performance outputs, recorded a process sigma value of 2.292.

The histogram and normal quantile plot of observed OPCi measurements are represented in Figure 3. The bell-shaped curve over the histogram clearly fits the distribution and is indicative of a normally distributed data. All measurements were within the lower and upper specification limits, with a little less than 55% of the providers' performance scores being above the mean of 12.76. The majority of the providers' mean scores ranging from 10.0 to 14.0 were recorded between the first and third quartiles. The normal probability and CDF plots of OPCi are displayed in Figure 4, with the data points falling approximately on the straight line, indicating

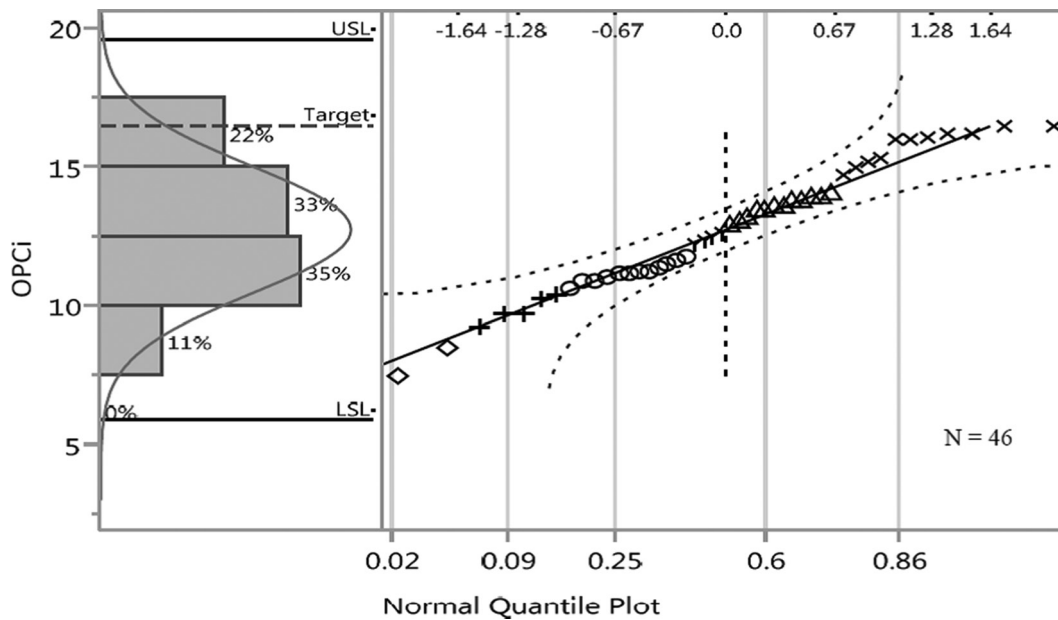
normality characteristics. There was a 50% probability of providers scoring below or above the mean of 12.76. The CDF plot indicates that there was 45% probability of HMCPs receiving OPCi scores that were between 12.76 (sample mean) and 16.50 (target mean).

Table 5 shows the capabilities indices of empirically-derived composite measure of preventive care counseling measurement system. On the basis of the process variability and specifications, our measurement system recorded capability indices of  $C_p = 0.999$  (95% CI: 0.793–1.205),  $C_{pk} = 0.999$  (95% CI: 0.771–1.227),  $C_{pm} = 0.238$ , and  $0.999$  (95% CI: 0.771–1.225) for  $C_{pu}$  and  $C_{pl}$ , respectively. All the measurements fell within the allowable process specifications, indicating that our current measurement system is statistically "stable" and "capable" (Figure 5). Expressing our measurement system capability in numerical term indicates that at long-term process sigma of 2.292 and  $C_{pk} = 0.999$ , the percentage of successful performance outputs was 72.81%. Approximately 27.19% of the measurements were captured outside allowable process limits leading to a nonconformance of 2,719.78 ppm at a sigma quality of 4.280. An attempt to shift the sample by  $1.5\sigma$  as part of the process control to improve the current performance toward the set mean target score of

**Table 4:** Descriptive statistics, specification limits<sup>a</sup> and prior probability, *Po* (out of control probability), and process sigma estimation of preventive care counseling performance

Statistics/specification limits	Value	Graphical representation	
LSL	5.89		
Target	16.50		
USL	19.63		
Mean	12.76		
STD	2.29		
CV (%)	17.96		
Mean shift normalized to specification	-0.272		
Standard deviation normalized to specification	0.167	<b>PPM</b>	<b>Sigma quality</b>
Below LSL (%)	13.58	1358.20	4.498
Above USL (%)	13.62	1361.58	4.497
Total outside (%)	27.20	2719.78	4.280
Prior probability, <i>Po</i>	0.272		

PPM (parts per million), the ppm value is the percent column multiplied by 10,000.  
<sup>a</sup>USL, upper specification limit; LSL, lower specification limit.



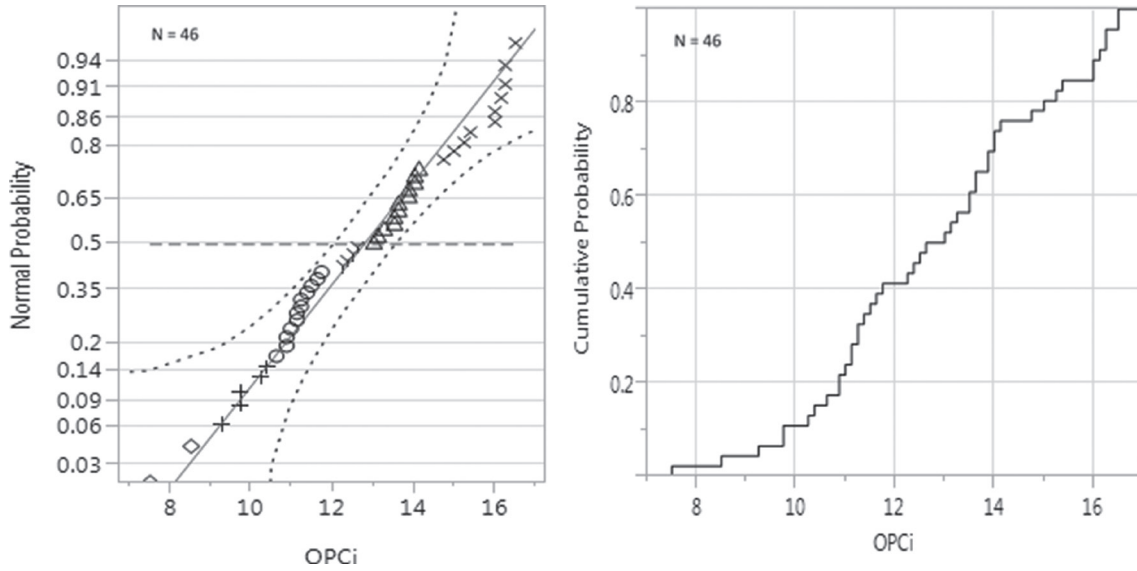
**Figure 3:** Histogram and normal quantile plot of overall preventive care counseling performance measurement.

16.50 (Figure 5) resulted in poor system outcomes. While the mean providers' performance improved by 26.94% to 16.198, the chance for nonconformance or not meeting the required specifications (outside allowable limits) increased by 146.9% to 6,715.40 ppm. In addition, a significant reduction in capability index ( $C_{pk} = 0.499$ ) following the  $1.5\sigma$  shift (system adjustment) implied that the system would be incapable of handling the performance measurements. Figure 6 presents the goal plots of how the OPCi data conformed to the normalized mean and standard deviation specification limits under normal

distribution at mean of 12.76 and process sigma of 2.292 at  $C_{pk} = 0.999$  and when shifted by  $1.5\sigma$  as part of the process control resulting in  $C_{pk} = 0.499$  and 67.15% nonconformance.

## Discussion

As with individual performance measure, the methods used to construct composites affect the reliability, validity, and usefulness of the composite measures and require some

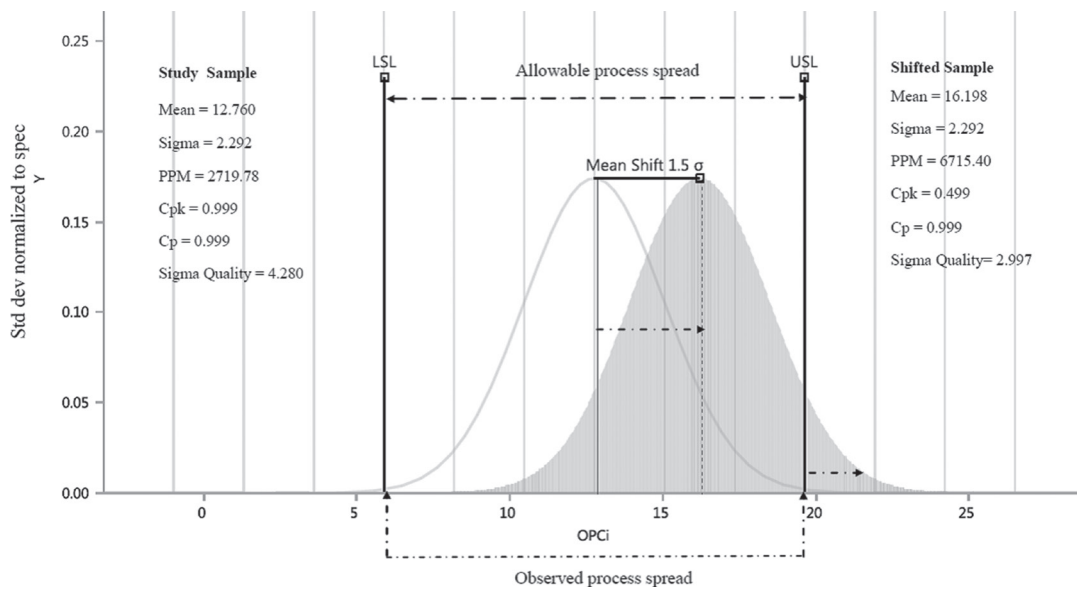


**Figure 4:** Empirical normal probability and cumulative distribution function (CDF) plots of OPCi.

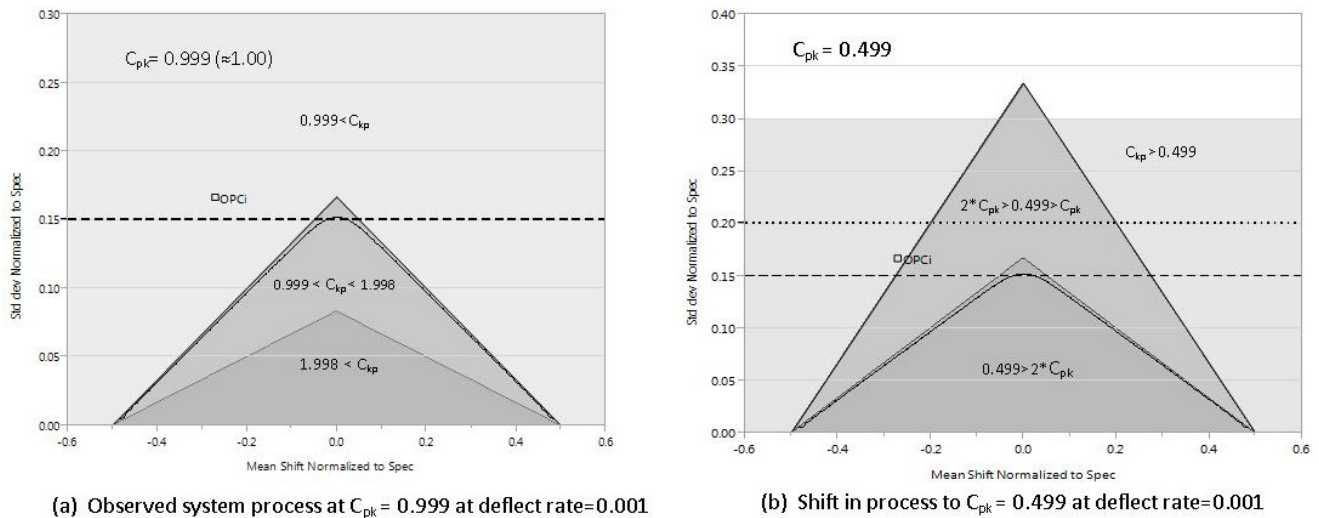
**Table 5:** Capabilities indices of empirically-derived composite measure of preventive care counseling measurement system

Capability measure	Index	Lower CI	Upper CI
Process capability ratio, $C_p$	0.999	0.793	1.205
Process capability index, $C_{pk}$	0.999	0.771	1.227
Process capability index, $C_{pm}$	0.238	— <sup>a</sup>	— <sup>a</sup>
Process capability ratio of one-sided upper spec, $C_{pu}$	0.999	0.771	1.225
Process capability ratio of one-sided lower spec, $C_{pl}$	0.999	0.771	1.225

<sup>a</sup> $C_{pm}$ , confidence interval (CI) not estimated because the target was not within the lower and upper spec limits range.



**Figure 5:** Preventive care counseling performance of HIV medical care providers: normal distribution based on process capability statistics from the study sample and shifted sample population.



**Figure 6:** Goal plots of OPCi with mean and standard deviation shift normalized to specifications ( $\Sigma = 2.292$ ).

unique considerations for testing and analysis.<sup>[12]</sup> The fairly high internal consistency reliability estimates obtained indicate that the indicators or items within the domain-specific composite indices measured the common underlying constructs. Our results support the finding that composite measures of performance tend to increase reliability.<sup>[12,16]</sup> Reliability is a key metric of the suitability of a measure for profiling because it describes how well one can confidently distinguish the performance of one provider from another<sup>[17]</sup> (i.e., the ability to detect a provider effect). However, it should be noted that reliability of a performance measure may be different than the rationale that all-or-none measures are intended to foster in a system perspective of care. In general, a set of measures used for profiling providers depend on how different the providers are from one another since reliability is not just a property of a measure but also depends on what population is used to estimate the reliability.<sup>[17]</sup> We believe that the composite measures of performance developed in our study would be useful in detecting a consistent pattern of practice or quality of care that reflects true performance difference among the HMCPs in Houston/Harris County, Texas, and elsewhere with similar environment.

The measurement system variation (gauge R&R) of 68.6% obtained in our study indicates the existence of significant systematic variation between and within HMCPs who evaluated the patients using the same preventive care counseling composite measures under different clinic settings. Furthermore, the contribution from gauge R&R was larger than that of the product (31.4%), implying that most of the variation was because of the measurement system—primarily repeatability (38.9%). This means that the measurement instrument cannot effectively decipher if the measure is good or bad, because too much measurement system variation showed up between the specifications. On the basis of the Barrentine guidelines

for measurement system variation,<sup>[30]</sup> our gauge R&R value was greater than 30% and, therefore, considered *unacceptable*. Although the main sources of variability are outside the performance measurement system's control because they are determined by the heterogeneity of the selected study population, identifying ways to reduce the variability across HMCPs' performance outputs and making them repeatable will help reduce variations noted in our current system. Doing so will require instituting intervention programs aimed at encouraging HMCPs to undertake continuing medical education and skills training on HIV prevention, be in compliance with standard guidelines, and enhance patients' education and communication. These efforts would augment the providers' performance outputs and the quality of preventive care received by HIV patients, leading to reduced variations across practices and subsequent improvement in our measurement system.

The low NDC value obtained in our study indicates that the measurement system cannot distinguish differences between the parts (in this case, newly diagnosed and established patients). It is possible that the noise of the measurement system obscured what differences there were between the parts. This assertion, however, contrasts our earlier finding in this study population, where significant difference in providers' preventive care counseling performance was noted between the patient groups in favor of newly diagnosed patients.<sup>[18]</sup> This happens because although a group of providers can measure all their patients equally well, some other providers tend to be biased towards a particular patient group; in this case, newly diagnosed patients; and raised questions about the practical use of NDC. Nevertheless, the Automobile Industry Action Group (AIAG) suggests that, if the number of distinct categories is five or more, it is considered a capable measurement system.<sup>[27]</sup> Similarly, our DRs were generally less than two, indicating

that the measurement system could not control the process (detect product variation). According to Wheeler,<sup>[24]</sup> DR above four is desirable. Unlike the NDC, the DR assumes that appraiser variation has been reduced to zero and carries the fractional part as well.<sup>[29]</sup> Given the discrete categories that are being used, the big question relates to the probability of misclassification. However, Wheeler reiterated that neither NDC nor DR defines the number of distinct categories or lend themselves to any practical interpretation.<sup>[29]</sup>

The high precision/tolerance ratios obtained for our measurement system represent the percent of the specification tolerance taken up by the measurement error. Our values of 99.13% for PTR and 82.85% for PTVR far exceed the 30% standard,<sup>[27,29,30]</sup> and, thus, made our measurement system *unacceptable*. This means that the measurement process cannot effectively decipher if the assessment is good or bad, because too much measurement system variation showed up between the specifications. Because gauge R&R accounts for a significant portion of the tolerance, it is possible that these measures may be more effective at detecting process mean shifts (for process control) than they are at identifying changes in process variation such as noted in this study. However, the acceptable level of precision depends on the context in which the measure is used. For instance, precision ratios may be made arbitrarily small by increasing the width of the specification tolerance.<sup>[30]</sup> Because these bounds are estimated based on the expected values, they may not be fully effective. In addition, these ratios provide no information about the capability of a measurement system to detect product variation. With an intraclass correlation, coefficient  $r_{pe}$  of 0.446 recorded for patient status in our study, the associated bias factor significantly reduced this proportion by 29.7% and represents a fraction of the measurement system variation attributed to reproducibility. According to Patki,<sup>[31]</sup> ignoring within-provider variation can mean overestimating the process variation, which in turn may result in tempering with a potentially satisfactory measurement system. He advised that great caution should be exercised in using these criteria to assess a measurement system in practice as it is easy to rule out a *good* measurement system as *unacceptable* due to nothing more than an oversight.

While some of the metrics come with certain recommended values that suggest the acceptability or unacceptability of the measurement system, most researchers<sup>[25–29,31]</sup> are querying the conditions under which the discrepancy occur and on which metrics, if any, are there relatively robust to variations in these conditions. Our study supports these notions because some of the metrics used to assess the current measurement system indicated that it is unacceptable and needs improvement, while our process capability study actually noted that the system expanded uncertainty was within set range at 99% confidence level making it statistical *stable* and *capable* at a long-term process sigma of 2.292 and  $C_{pk} = 0.999$  with successful performance outputs of 72.81% and 27.19% nonconformance. While we agree that our measurement system needs some improvement, it seems, however, that

the true test of Gauge R&R is determined as “a judgment call” depending on the process, intended objectives, cost, and other environmental factors associated with the measurement.

In attempt to shift the sample mean by  $1.5\sigma$  (composite score = 16.50) as part of the process control measure to improve the current provider performance, we noted that the system tends to perform poorly. With a modest mean increase of 26.94%, the system witnessed large increase (146.9%) in nonconformance from 2,719.78 to 6,715.40 ppm with a significant reduction in process capability index ( $C_{pk} = 0.499$ ) and sigma quality of 2.997. This implied that, under this condition, the system would be incapable of handling the performance measurements. Consequently, as the providers under study increase their performance, the reliability may decrease if the provider-to-provider variance decreases over time.<sup>[17]</sup> This is especially true as the measure hit the upper limits of the ranges. It is also possible for the degree of measurement error to be overstated if the process spread is not fully represented.<sup>[32]</sup> Reliable processes are a proven way to better outcomes. Consequently, it is necessary to improve the processes first to make them extremely reliable before considering outcomes improvements.<sup>[17]</sup> This may essentially require adopting a balancing act for both process and outcome measures in order to attain set goals. But, for this approach to work, the data must be trustworthy, and the precision and accuracy of the measurement system should be satisfactory and, hence, quantifiable.<sup>[28,32]</sup> An ideal measurement system produces measurements that agree exactly with a master standard (e.g., target goal). Unfortunately, measurement systems with such properties rarely exist. Conceptual challenges remain, especially as multiple levels in the system may influence measures score.

### Study Limitations and Strength

The HMCPs who responded to the survey made up a relatively small sample (45%), and may not be fully representative of the providers in Houston/Harris County, Texas. Although probability proportional to size sampling method was used to select participating facilities, it was not possible to weight the providers' self-reported responses to the survey questions because of the small sample size, and, therefore, some performance estimates may be subject to nonresponse bias. Our data was based on self-report by providers and was not confirmed through patient interviews or clinical records and, thus, are also subject to recall and social desirability bias. It is possible that prevention counseling delivery was over- or under-reported. The differential weights applied to the composite prevention counseling indices were based on conceptual rationale<sup>[21,22]</sup> and not on the comparative importance of the care process or counseling quality and contents.<sup>[18]</sup> In addition, the HMCPs' practices and awareness may have evolved since the data were collected in 2009.

Despite these limitations, the strength of this study lies in being the first of its kind to evaluate the measurement system of empirically-derived composite measures of preventive care counseling practices of HMCPs. Information obtained

from this study lays a strong foundation for the assessment of providers' performance in preventive care and the measurement process, which are not typically captured through routine public health surveillance activities. Our measurement system was determined to be statistically *stable* and *capable*, and with little improvement in the process delivery, it can be effectively used to assess providers' preventive care efforts in HIV clinical settings, especially in the light of the current emphasis on HIV prevention,<sup>[6-8]</sup> and pay-for-performance in the new health care delivery system in the United States and some other countries.

### Research Implications

As the nation moves forward, improving its healthcare systems, performance measurement will be fundamental to understanding healthcare's current state and critical to developing solutions for improvement.<sup>[12]</sup> According to Institute of Medicine (IOM) report, composite measures can enhance measurement to extend beyond tracking performance on separate measures and can provide a potentially deeper view of the reliability of the care system.<sup>[17]</sup> In addition, this approach can support efforts to rate providers by the quality of their care, while bringing about the need for better preventive care coordination and accountability. Our study identified the existence of large measurement variations (Gauge R&R = 68.6%), which significantly affected the outcome measure. Differences between providers and their individual measurements are largely outside the performance measurement system's control because they are determined by the heterogeneity of the study population. This finding corroborates with the low and inconsistent preventive care counseling practices reported in an earlier study with this population.<sup>[18]</sup>

Measurement systems generally tend to degrade over time.<sup>[17,29]</sup> Thus, it is important to regularly assess measurement systems to validate that they continue to provide reliable data—whether those systems are people's judgment calls or measuring instruments or processes. The ultimate goal of our assessment was to eliminate as much analytical variability as possible, thereby reducing the contribution of nonnatural effects on the reported outcome. An ineffective measurement system can have a dramatic impact on provider performance especially as it can lead to uninformed (and usually bad) decision-making that can mislead patients, managed care, and monitoring organizations. Defining appropriate control strategies and further testing of the system processes for stability and capability will be important for the widespread dissemination of the current measurement system to support HIV prevention efforts in medical settings. It is, therefore, more cost effective to improve the measurement system by reducing the measurement variations than to embark on another major process or system design changes. Such efforts could enhance both the accuracy of medical care providers' performance outputs and the quality of preventive care received by HIV patients, leading to reduced variations across practices and improvement in the current measurement system.

### Conclusions

MSA is a key step to any process improvement effort. Having identified the sources of variations in the measurement system, we are in a better position to address the issues identified and make better recommendations for its improvement. Providers should have enough information to guarantee that HIV patients are not treated differently under the same context, and they should have similar performance scores under similar clinical condition or environment. In an ideal situation, a single measurement model should be applicable to a given provider and indeed the whole healthcare system. The process control objective will be to have each provider comply strictly with recommended guidelines and procedures for HIV prevention and to meet the performance target regardless of the patient-provider combinations by eliminating special causes of variation so that only common (natural) causes are acting on the process and to reduce those to barest minimum, whenever possible. Understanding exactly what information is needed for effective process control will allow for better planning and integration of HIV preventive care services with overall clinical practice needs and strategies.

However, with HMCPs' process and practice environment being markedly different from each other; the greatest challenge will be how to minimize the variations associated with these factors in order to enhance the current measurement system and performance output of providers as evident in the quality of preventive care received by HIV patients.<sup>[18]</sup> A system combining all of the appropriate preventive care measures such as identified in our study can help the patient become a more informed and an active participant in their care. Future research should explore the extent to which empirically-derived composite measures or constructs can provide stable performance rankings of providers over time and to assess whether variations are due to genuine performance improvement or merely the result of random statistical variations. Addressing variation in quality of HIV preventive care offered to patients by providers is perhaps at the heart of any efforts to improve patient care outcomes. Ultimately, it would also be necessary for future research to target the dynamic relationships that exist between medical practice structure, measurement system, patient care outcome, and patient experience.

### Acknowledgment

The authors would like to thank the HIV medical care providers who participated in this survey, the MMP field staff and members of the Community Advisory Board and Provider Advisory Board in Houston project area; the Houston Department of Health and Human Services and members of the Clinical Outcomes Team in CDC's Behavioral and Clinical Surveillance Branch of the Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. They also thank the reviewers for their very insightful comments.

## References

1. Erbeling EJ, Stanton D, Quinn TC, Rompalo A. Behavioral and biologic evidence of persistent high-risk behavior in an HIV primary care population. *AIDS* 2000;14:297–301.
2. Catz SL, Meredith KL, Mundy LM. Women's HIV transmission risk perceptions and behaviors in the era of potent antiretroviral therapies. *AIDS Educ Prev* 2001;13:239–51.
3. Marks G, Crepaz N. HIV-positive men's sexual practices in the context of self-disclosure of HIV status. *J Acquir Immune Defic Syndr* 2001;27:79–85.
4. Crepaz N, Marks G. Toward an understanding of sexual risk behavior in people living with HIV: a review of social, psychological, and medical findings. *AIDS* 2002;16:135–149.
5. Centers for Disease Control and Prevention. Incorporating HIV prevention into the medical care of persons living with HIV. *MMWR Morb Mortal Wkly Rep*. 2003;52(RR-12):1–24.
6. Centers for Disease Control and Prevention. Advancing HIV prevention: new strategies for a changing epidemic—United States, 2003. *MMWR Morb Mortal Wkly Rep* 2003;52:329–32.
7. White House Office of National AIDS Policy. *National HIV/AIDS Strategy for the United States*. July 2010. Available at: <http://aids.gov/federal-resources/national-hiv-aids-strategy/nhas.pdf> (last accessed on May 10, 2014).
8. Centers for Disease Control and Prevention. Vital signs: HIV prevention through care and treatment—United States. *MMWR Morb Mortal Wkly Rep* 2011;60(47):1618–23.
9. James J. Health policy brief: pay-for-performance. *Health Affairs*. October 11, 2012. pp. 1–6. Available at: [www.healthaffairs.org](http://www.healthaffairs.org) (Last accessed on August 10, 2014).
10. *Performance Measurement: Accelerating Improvement*. Committee on Redesigning Health Insurance Performance Measures, Payment, and Performance Improvement Programs. Institute of Medicine. Washington, DC: National Academies Press, 2006.
11. American Medical Association. *Measures Development, Methodology, and Oversight Advisory Committee: Recommendations on Composite Measures*. Physician Consortium for Performance Improvement (PCPI). Chicago, IL: American Medical Association, December 2010.
12. National Quality Forum (NQF). *Composite Measure Evaluation Framework and National Voluntary Consensus Standards for Mortality and Safety—Composite Measures: A Consensus Report*. Washington, DC: NQF, 2009.
13. Zaslavsky AM, Shaul JA, Zaboriski LB, Cioffi MJ, Cleary PD. Combining health plan performance indicators into simpler composite measures. *Health Care Financ Rev* 2002;23:101–15.
14. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care* 2007;45:489–96.
15. Caidis T. Composite health plan quality scales. *Health Care Financ Rev* 2007;28:95–107.
16. Kaplan SH, Normand ST. *Conceptual and Analytical Issues in Creating Composite Measures of Ambulatory Care Performance*. Washington, DC: NQF, December 2006.
17. Adams JL. *The Reliability of Provider Profiling: A Tutorial*. National Committee for Quality Assurance Technical Report. Santa Monica, CA: Rand Health, 2009. 33p.
18. Mgbere O, Rodriguez-Barradas MC, Bell TK, Khuwaja S, Arafat R, Essien EJ, et al. Frequency and determinants of preventive care counseling by HIV medical care providers during encounters with newly diagnosed and established HIV patients. *J Int Assoc Provid AIDS Care* 2014. doi: 10.1177/2325957414556352 [Epub ahead of print].
19. McNaghten AD, Valverde EE, Blair JM, Johnson CH, Freedman MS, Sullivan PS. Routine HIV testing among providers of HIV care in the United States, 2009. *PLoS One*, 2013;8(1):e51231.
20. Mgbere O, Khuwaja S, Bell T, Rodriguez-Barradas M, Arafat R, Blair J, et al. Managing the personal side of health care among HIV/AIDS patients: a pilot study of providers' perspectives. *J Int Assoc Provid AIDS Care* 2014. pii: 2325957414555229 [Epub ahead of print].
21. Smith P. Developing composite indicators for assessing health system efficiency. In: *Measuring Up: Improving the Performance of Health Systems in OECD countries*, Smith PC (Ed.). Paris: OECD, 2002.
22. Freudenberg M. *Composite Indicators of Country Performance: A Critical Assessment*. OECD STI Working paper DSTI/DOC 2003/16. Paris: OECD, 2003.
23. *JMP Version 11: Quality and Process Methods*. Cary, NC: SAS Institute, Inc., 2013. p. 260.
24. Wheeler DJ. *EMP III (Evaluating the Measurement Process): Using Imperfect Data*. Knoxville, TN: SPC Press, 2006. 316 pp.
25. Montgomery DC, Runger, GC. Gauge capability analysis and designed experiments. Part I: basic methods. *Qual Eng* 1993;6(1):115–35.
26. Montgomery DC, Runger, GC. Gauge capability analysis and designed experiments. Part II: experimental design models and variance component estimation. *Qual Eng* 1993;6(2):289–305.
27. *AIAG Measurement Systems Analysis, Reference Manual, 3rd ed.* Southfield, MI: Automotive Industry Action Group, 2003. Available at: [www.aiag.org](http://www.aiag.org). (Last accessed on September 29, 2014).
28. Wheeler DJ, Lyday RW. *Evaluating the Measurement Process*. Knoxville, TN: SPC Press, Inc., 1984.
29. Wheeler DJ. An Honest Gauge R&R Study. *2006 ASQ/ASA Fall Technical Conference*. Manuscript No. 189. Published in SPC Ink March 2008, revised January 2009. Available at: [www.spcpress.com/pdf/DJW189.pdf](http://www.spcpress.com/pdf/DJW189.pdf). (Last accessed on October 18, 2014).
30. Barrentine LB. *Concepts for R&R Studies*. Milwaukee, WI: ASQC Quality Press, 1991.
31. Patki M. *Investigation, Improvement and Extension of Techniques for Measurement System Analysis*. Doctor of Philosophy Dissertation, Faculty of the Graduate College of Oklahoma State University, 2005, 150 pp.
32. *Evaluating the Measurement Systems (EMP Studies) and Measurement System Analysis*. Six Sigma Study Guide. Available at: <http://www.sixsigmastudyguide.com/evaluating-the-measurement-systems-emp-studies/> (last accessed on July 12, 2014).

**How to cite this article:** Mgbere O, Singh M, Arafat R. Measurement system analysis (MSA) of empirically-derived composite measure of preventive care counseling practices of HIV medical care providers. *Int J Med Sci Public Health* 2015;4:1640-1654

**Source of Support:** The Medical Monitoring Project (MMP) 2009 Provider Survey was supported by the Centers for Disease Control and Prevention (CDC) under cooperative agreement number PS09-937. The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention. **Conflict of Interest:** None declared.